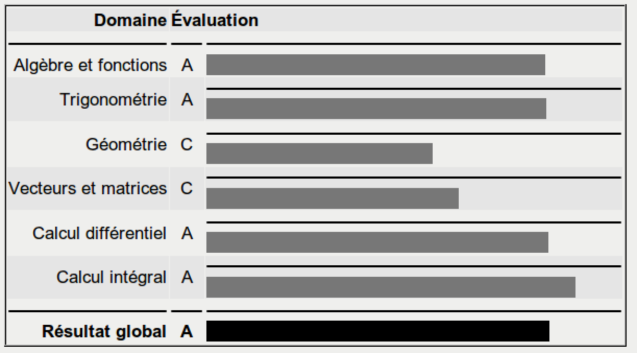
Conditions for effectively deriving a Q-Matrix from data with Non-negative Matrix Factorization

Michel C. Desmarais

Polytechnique Montréal

Educational Data Mining, EDM2011, Eindhoven
July 8, 2011

Need to score a number of topics from a set of question items?



Scoring an exam

	Results	Topic
Question 1	1	X
Question 2	0	X
Question 3	1	Y
Question 4	1	Z
Question 5	1	Z
Question 6	0	Z

Topic mastery:

Topic X: 1/2 — Topic Y: 1/1 — Topic Z: 2/3

Scoring an exam

	Results	Topic X	Topic Y	Topic Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrix

Topic mastery:

Topic X: 1/2 — Topic Y: 1/1 — Topic Z: 2/3

Scoring an exam

	Results	Topic X	Topic Y	Topic Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrix

Topic mastery:

Topic X: 1/2 — Topic Y: 1/1 — Topic Z: 2/3 $\Rightarrow$ $[\frac{1}{2}\frac{1}{1}\frac{2}{3}]^T = \mathbf{Q}^T[101110]^T$

Scoring an exam

	Results	Topic X	Topic Y	Topic Z
Question 1	1	1	0	0
Question 2	0	1	0	0
Question 3	1	0	1	0
Question 4	1	0	0	1
Question 5	1	0	0	1
Question 6	0	0	0	1

Q-matrix

Topic mastery:
Topic X: 1/2 — Topic Y: 1/1 — Topic Z: 2/3 $\Rightarrow$ $[\frac{1}{2}\frac{1}{1}\frac{2}{3}]^T = \mathbf{Q}^T[101110]^T$

(1) $\mathbf{S} = \mathbf{Q}^T \cdot \mathbf{R}$ (skill mastery, $\mathbf{S}$, is the dot product $\mathbf{Q}$ and results matrix $\mathbf{R}$—assuming column normalization)

(2) $\mathbf{R} = \mathbf{Q} \cdot \mathbf{S}$

Q-matrix basics

$$\begin{array}{l} \\ \text{Question 1} \\ \text{Question 2} \\ \text{Question 3} \\ \text{Question 4} \\ \text{Question 5} \\ \text{Question 6} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Topic X} & \text{Topic Y} & \text{Topic Z} \end{array} \\ \boxed{\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}} \end{array} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\end{array}}\right\} \textbf{Q-matrix}$$

Flavours of Q-matrices:

- Can have One or many skills per question item
- Conjunctive models: all skills necessary for success
- Disjunctive: any skills can lead to success
- Compensatory: each skill adds to the chances of success

Mining data to derive the Q-matrix

Objective: derive Q-matrix from data

- Early work by Tatsuoka (1983) and a recent book on the topic (2009)
- Algorithm developed by Barnes (2006) and compared to PCA clustering
- For dynamically changing knowledge states, Cen et al. (2005, 2006) have used LFA
- Thai-Nghe, Horváth and Schmidt-Thieme (EDM 2011)
- Current study focuses on static knowledge states and **Non-negative Matrix Factorization (NMF)**
- Follows up on the work of **Winters et al. (2005)**

Non-negative Matrix Factorization (NMF)

Results can be obtained by the product of the Q-matrix with the skills mastery matrix
$\mathbf{R} = \mathbf{Q}\ \mathbf{S}$ (transformation of $\mathbf{S} = \mathbf{Q}^T\ \mathbf{R}$)
In the standard NMF notation, this is:
$\mathbf{X} = \mathbf{W}\ \mathbf{H}$

Characteristics of NMF:

- Mapping of $\mathbf{W} \to \mathbf{Q}$, $\mathbf{H} \to \mathbf{S}$, and $\mathbf{X} \to \mathbf{R}$
- Matrices $\mathbf{W}$ and $\mathbf{H}$ are forced to be positive
- Natural interpretation in terms of a **compensatory model**

Comparison

Numerous techniques to matrix factorization with similar aims:

- **PCA**: Principal Component Analysis: similar to **NMF** but imposes orthogonality on the eigenvector matrix (Q-matrix equivalent in our case).
- The **orthogonality** constraint generally implies an optimized and unique solution, but it imposes vector independence and negative values

NMF advantages and issues

NMF advantages:

- Non orthogonality (independence) of skills in Q-matrix equivalent is more realistic compared to PCA
- Natural interpretation in terms of a compensatory model and positive values for skills
- Many standard algorithms available

NMF issues:

- Multiple solutions
- Non determinism of some algorithms
- Number of skills determined *a priori* (not specific to NMF)
- **Does it work and is the result interpretable? At least under restricted cases.**

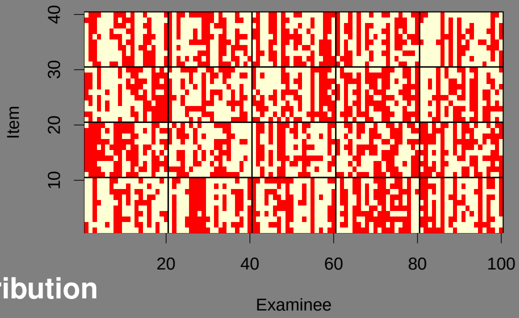
Experiment

Simulated Data

Does NMF work under restricted cases?

Simulated data

Results $(\mathbf{R} = \mathbf{QS})$
Results $(\mathbf{X} = \mathbf{WH})$
4 skills
40 question items
100 examinees
No correlation between skills
Random normal (0,1) distribution

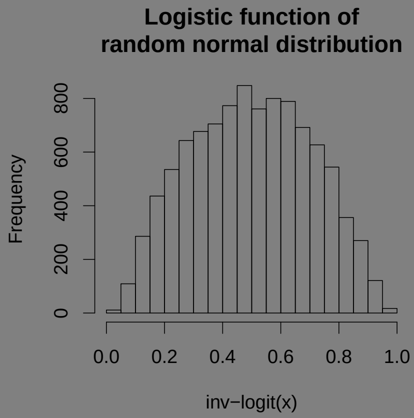


Sampling procedure: each column contains 40 probabilities, one probability per item, structured as a sequence of 10×4 probabilities:

$$(p_{1,1}, p_{1,2}, ..., p_{1,10}, p_{2,1}, ..., p_{2,10}, p_{3,1}, ..., p_{3,10}, p_{4,1}, ..., p_{4,10})$$

where $p_{1,1}$ to $p_{1,10}$ are all equal, $p_{2,1}$ to $p_{2,10}$ are all equal, and so on

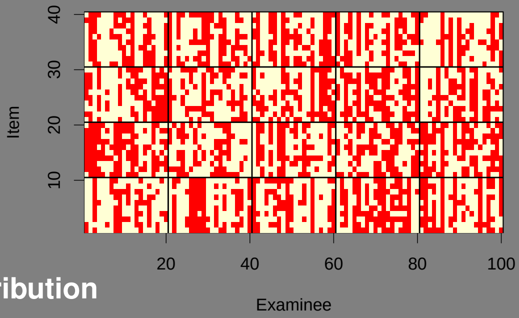
Logistic function of random normal distribution



Does NMF work under restricted cases?

Simulated data

Results ($\mathbf{R} = \mathbf{QS}$)
Results ($\mathbf{X} = \mathbf{WH}$)
4 skills
40 question items
100 examinees
No correlation between skills
Random normal (0,1) distribution



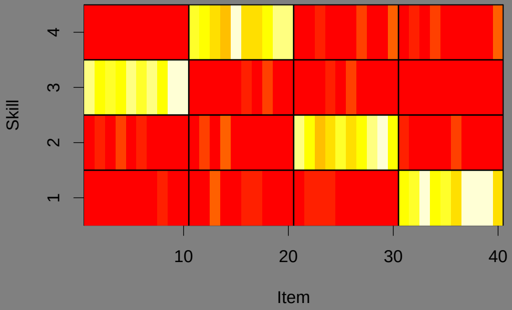
Sampling procedure: each column contains 40 probabilities, one probability per item, structured as a sequence of 10×4 probabilities:

$$(p_{1,1}, p_{1,2}, \ldots, p_{1,10}, p_{2,1}, \ldots, p_{2,10}, p_{3,1}, \ldots, p_{3,10}, p_{4,1}, \ldots, p_{4,10})$$

where $p_{1,1}$ to $p_{1,10}$ are all equal, $p_{2,1}$ to $p_{2,10}$ are all equal, and so on

Q-matrix from simulated data

Q-Matrix $(\mathbf{R} = \mathbf{QS})$
Q-Matrix $(\mathbf{X} = \mathbf{WH})$
4 skills
40 question items
$\Rightarrow$ **Perfect classification!**

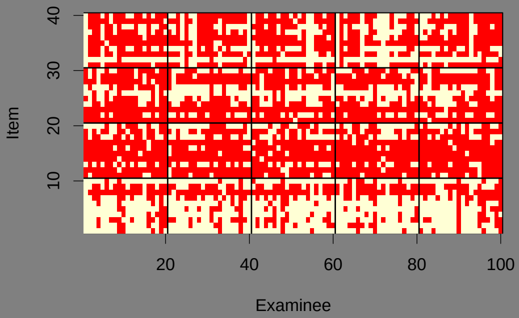


Experiment

Real Data

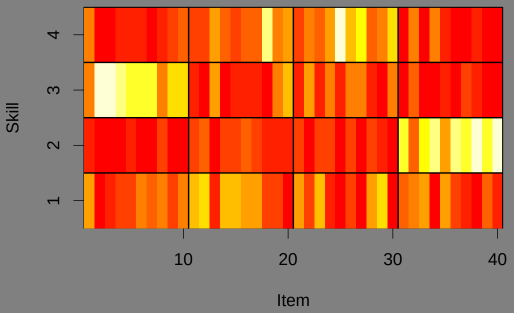
Real data on SAT test score

Results $(\mathbf{X} = \mathbf{WH})$
SAT: Math, Biology, World history and French
40 question items



Real data on SAT test score

Q-Matrix ($\mathbf{X} = \mathbf{WH}$)
SAT: (1) Math, (2) Biology, (3) World history and (4) French
40 question items
$\Rightarrow$ Accurate only for math and French

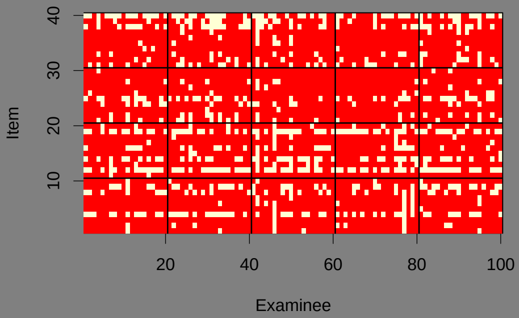


Real data on Trivia test scores

Trivia: (1) Arts and entertainment, (2) Sports and leisure, (3) Science and nature, and (4) Geography

40 question items

Note the low success rate

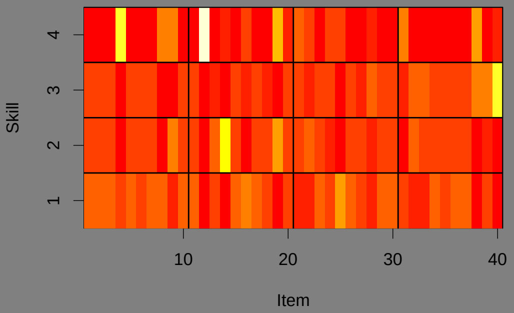


Real data on Trivia test scores

Trivia: (1) Arts and entertainment, (2) Sports and leisure, (3) Science and nature, and (4) Geography

40 question items

⇒ **No obvious pattern**



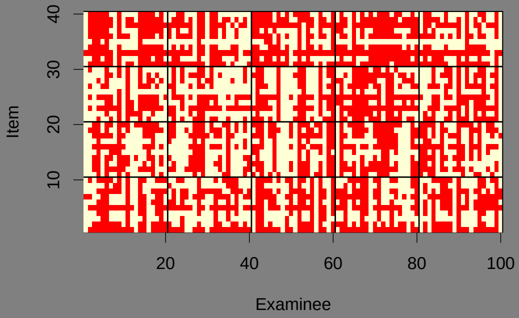
Experiment

More realistic simulated data

Realistic simulated data

Addition of item difficulty and examinee ability

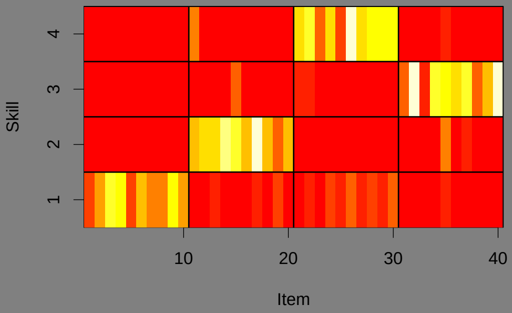
Model with skills, item difficulty and examinee ability:
$P(X_{mnq}) = \Phi(\beta_m + \beta_n + \beta_q)$



Realistic simulated data

Addition of item difficulty and examinee ability

Model with skills,
item difficulty and
examinee ability
⇒ **Also a 100% accuracy**,
as found with skills only
model



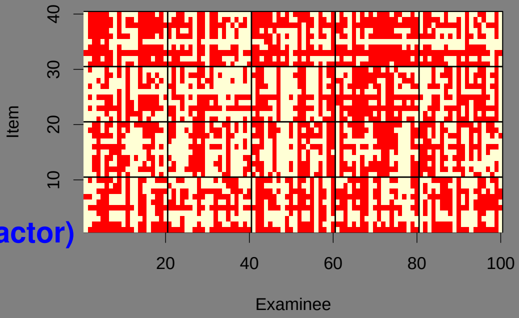
Exploring the space of parameters

Realistic simulated data

Addition of item difficulty and examinee ability

Model with skills, item difficulty and examinee ability:
$P(X_{mnq}) = \Phi(\beta_m + \beta_n + \beta_q)$

Let us vary β_q(topic skill factor)



Parameter Space

	Topic skill (β_q)		Accuracy	
	Mean	S.d.	Mean acc.	S.d. acc.
1	0	0	0.36	0.05
2	0	0.10	0.48	0.07
3	0	0.25	0.60	0.11
4	0	0.50	0.93	0.08
5	0	1	1	0
Trivia data parameters				
6	0	0.12	0.75	0.12
7	*n.a.*	*0.12*	*0.35*	*0.03*
SAT data parameters				
8	0	0.24	0.98	0.05
9	*n.a.*	*0.24*	*0.72*	*0.02*
10	*n.a.*	*0.24*	*0.96*	*0.05*

Parameter Space

	Topic skill (β_q)		Accuracy	
	Mean	S.d.	Mean acc.	S.d. acc.
1	0	0	0.36	0.05
2	0	0.10	0.48	0.07
3	0	0.25	0.60	0.11
4	0	0.50	0.93	0.08
5	0	1	1	0
Trivia data parameters				
6	0	0.12	0.75	0.12
7	*n.a.*	*0.12*	*0.35*	*0.03*
SAT data parameters				
8	0	0.24	0.98	0.05
9	*n.a.*	*0.24*	*0.72*	*0.02*
10	*n.a.*	*0.24*	*0.96*	*0.05*

Parameter Space

	Topic skill (β_q)		Accuracy	
	Mean	S.d.	Mean acc.	S.d. acc.
1	0	0	0.36	0.05
2	0	0.10	0.48	0.07
3	0	0.25	0.60	0.11
4	0	0.50	0.93	0.08
5	0	1	1	0
Trivia data parameters				
6	0	0.12	0.75	0.12
7	*n.a.*	*0.12*	*0.35*	*0.03*
SAT data parameters				
8	0	0.24	0.98	0.05
9	*n.a.*	*0.24*	*0.72*	*0.02*
10	*n.a.*	*0.24*	*0.96*	*0.05*

Parameter Space

	Topic skill (β_q)		Accuracy	
	Mean	S.d.	Mean acc.	S.d. acc.
1	0	0	0.36	0.05
2	0	0.10	0.48	0.07
3	0	0.25	0.60	0.11
4	0	0.50	0.93	0.08
5	0	1	1	0
Trivia data parameters				
6	0	0.12	0.75	0.12
7	*n.a.*	*0.12*	*0.35*	*0.03*
SAT data parameters				
8	0	0.24	0.98	0.05
9	*n.a.*	*0.24*	*0.72*	*0.02*
10	*n.a.*	*0.24*	*0.96*	*0.05*

Discussion

Discussion

- Under assumptions of a **simple model** ($P(X_{mnq}) = \Phi(\beta_m + \beta_n + \beta_q)$), NMF is highly effective for deriving the Q-matrix structure as shown with simulated data
- In reality, NMF can:
  - correctly assign **Math. vs. French** questions; however, this is not highly challenging...
  - cannot discern **Trivia** questions over different topics

Discussion

- Under assumptions of a **simple model** ($P(X_{mnq}) = \Phi(\beta_m + \beta_n + \beta_q)$), NMF is highly effective for deriving the Q-matrix structure as shown with simulated data
- In reality, NMF can:
  - correctly assign **Math. vs. French** questions; however, this is not highly challenging...
  - cannot discern **Trivia** questions over different topics

Discussion

- We conclude that:
  - **topic skills are sometimes insignificant**
  - **the simple model does not reliably reflect real data**

    $\Rightarrow$ we have a **gap** in simulation results with synthetic data
- The **methodology** may be useful to quantify the topic skill effect as it provides an estimate of:
  - the expected/theoretical **skill effect** under simple model
  - the **gap** from the simple model to the reality
  - and therefore, some idea of **model fit**

Parameter Space

	Topic skill (β_q)		Accuracy	
	Mean	S.d.	Mean acc.	S.d. acc.
1	0	0	0.36	0.05
2	0	0.10	0.48	0.07
3	0	0.25	0.60	0.11
4	0	0.50	0.93	0.08
5	0	1	1	0
Trivia data parameters				
6	0	0.12	0.75	0.12
7	*n.a.*	*0.12*	*0.35*	*0.03*
SAT data parameters				
8	0	0.24	0.98	0.05
9	*n.a.*	*0.24*	*0.72*	*0.02*
10	*n.a.*	*0.24*	*0.96*	*0.05*

Discussion

- We conclude that:
  - **topic skills are sometimes insignificant**
  - **the simple model does not reliably reflect real data**

    $\Rightarrow$ we have a **gap** in simulation results with synthetic data
- The **methodology** may be useful to quantify the topic skill effect as it provides an estimate of:
  - the expected/theoretical **skill effect** under simple model
  - the **gap** from the simple model to the reality
  - and therefore, some idea of **model fit**

Discussion

- We conclude that:
  - **topic skills are sometimes insignificant**
  - **the simple model does not reliably reflect real data**

    $\Rightarrow$ we have a **gap** in simulation results with synthetic data
- The **methodology** may be useful to quantify the topic skill effect as it provides an estimate of:
  - the expected/theoretical **skill effect** under simple model
  - the **gap** from the simple model to the reality
  - and therefore, some idea of **model fit**

Exploring the space of parameters

Unix commands data set and the number of skills

Exploring the number of skills

Unix commands

Q-01 base	Q-11 print	Q-24 io
Q-02 base	Q-12 print	Q-25 io
Q-03 base	Q-13 print	Q-26 io
Q-04 base	Q-14 dir	Q-27 shell
Q-05 base	Q-15 dir	Q-28 regexp
Q-06 base	Q-16 dir	Q-29 regexp
Q-07 base	Q-17 admin	Q-30 regexp
Q-08 shell	Q-18 admin	Q-31 regexp
Q-09 shell	Q-19 admin	Q-32 awk
Q-10 shell	Q-20 fpermis	Q-33 awk
	Q-21 fpermis	Q-34 awk
	Q-22 fpermis	
	Q-23 fpermis	

Examples:

Q1: `ls`

Q4: `cat`

Q8: `which`

Q12: `lpq`

Q18: `/etc/mtab`

Q22: `-rwxr-xr-x`

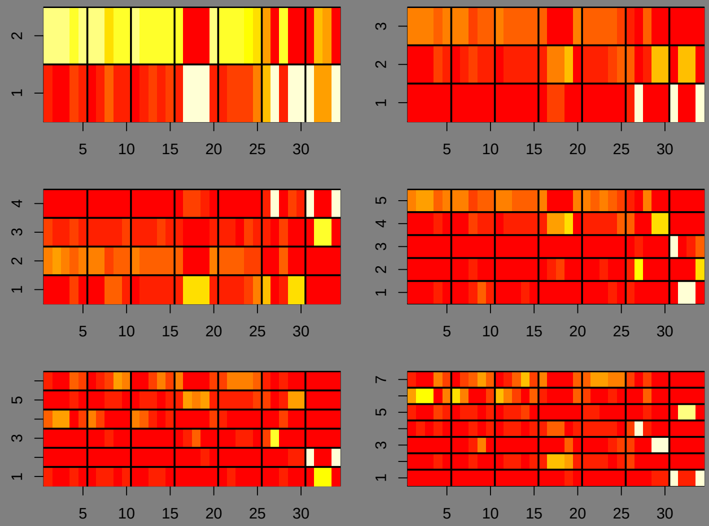
Q25: `ls | sort > b`

Q29: `s/b/x/`

Q32: `{print $2 + $3}`

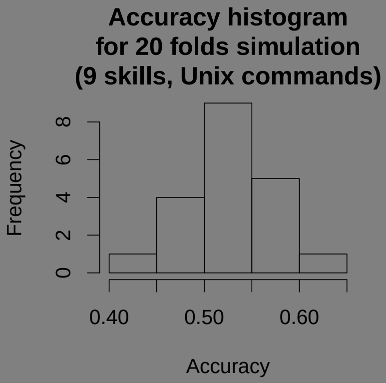
Exploring the number of skills

Unix commands with various number of skills



Simulations with 9 skills of Unix commands

Accuracy performance



34 items; 48 respondants

20 folds NMF simulations

Accuracy for random: 0.301

Maximum accuracy unknown (comparison with an expert classification)